[SMU Data Science Review](https://scholar.smu.edu/datasciencereview)

Volume 6
Number 2 *Summer 2022*

Article 16

2022

A Framework for Predicting the Optimal Price and Time to Sell a Home

Adeel Qureshi
Southern Methodist University, AdeelQ@smu.edu

Iosif Mushailov
Southern Methodist University, IMushailov@smu.edu

Patricia Herrera
Southern Methodist University, triciah@smu.edu

Phillip Hale
Southern Methodist University, PhilipH@smu.edu

Reannan McDaniel
rmcdaniel419@gmail.com

A Framework for Predicting the Optimal Price and Time to Sell a Home

Patricia Herrera [1], Iosif Mushailov [2], Adeel Qureshi [3] Phillip Hale [4] and Reannan McDaniel [5]
Master of Science in Data Science, Southern Methodist University, Dallas, TX 75275 USA
{TriciaH, IMushailov, AdeelQ, PhilipH}smu.edu and rmcdaniel419@gmail.com

Abstract. Due to high barriers to conduct housing market research, many home sellers opt to go to the market with asymmetric information or invest large sums of money into hiring a professional. This research aims to reduce these inefficiencies by proposing a framework that provides sellers with a concrete recommendation on optimal time and price to sell a home to maximize financial gains. The core data used in this research is the NOVA Home Price dataset, which contains 34,973 house listings over multiple years in Northern Virginia. A pipeline of machine learning models, including a linear regression, random forest, XGboost and artificial neural network are trained and evaluated for performance on predicting home close prices. The final model employed is an ensemble of random forest and XGboost and is tested on both a holdout set of Northern Virginia data as well as real estate data scraped from Zillow to introduce some variance. To control for future economic trends, a long-short-term memory model is then trained using temporal data from the Federal Reserve. Finally, the algorithm distills the insights from the disparate models to provide recommendations on optimal time and price to go to market, as well as short-term investments to increase potential gains from sale. The study finds that home features coupled with macro-economic trends can offer home sellers strong recommendations on optimal time and price to list homes. This research is preliminary and should be used as a baseline for future studies.

1 Introduction

Predicting home values is one of the most sought-after areas of real estate research. This interest has resulted in the development of home value estimator tools such as Zillow, Redfin, Trulia, Homelight, and ReMax to help home sellers value and sell their homes. A seller's decision to use one of the numerous home estimator tools is just one of the many decisions that must be made when selling a home. Endogenous factors such as quality of appliances, window fixtures, driveway quality (and, for example, whether repaving prior to sale is a good investment), lighting fixtures, siding, roofing, etc. can all play important roles in the value of a home. Additionally, exogenous factors like time of year and economic indicators also influence home prices. The many decisions a seller must make can be overwhelming, which is why so many seek the assistance of a realtor or appraiser. This research aims to develop a framework to help sellers make more informed decisions about what price to sell homes at, and when, without relying on a paid professional, or dedicating extensive time and money into market research.

Google Trends shows a sharp rise in searches for "Zillow" starting in 2010, continuing into 2016, followed by moderate increases in the search term (Trends, 2022). Zestimate was the first product Zillow launched when the company was founded in 2006, according to now Chief Analytics Officer/Economist at Zillow, Stan Humphries (Schneider, 2019). These automated appraisals were meant to offer an unbiased of home prices, as homeowners have been shown to overestimate home value (John et al., 1992). The growth of the machine learning field has enabled home price evaluation methodology to become more complex and accurate but utilizing traditional statistical techniques like regression to appraise homes is nothing new (Benjamin et al., 2004). However, despite their ubiquity, these tools have one major gap: they do not provide clear recommendation on how to maximize home seller profits while adjusting for both exogenous and endogenous factors in the markets.

According to Rockett, a leading home loan provider, as of January 2022 some important exogenous factors directly affecting home selling are inflation, US housing inventory and mortgage interest rates (Rockett, 2022). Both inflation and inventory are addressed in this preliminary research, but this study does not cover mortgage interest rates, price stickiness, and affordability tiers due to the complexity of the topic. The objective of this study is to establish a framework the impact of housing inventory, inflation, and home features on optimal time and price to sell a home, with additional research to follow.

A key economic factor affecting home prices is inflation. It has a standard definition as a continuing rise in the general price level usually attributed to an increase in the volume of money and credit relative to available goods and services (Merriam Webster, 2022). Inflation plays a part in the house price by raising the cost of housing materials, increasing market price, and decreasing the buying power of home buyers. The inflation rate in the United States has been relatively stable over the last 20 years up until the COVID-19 pandemic (Belz et al., 2020). According to the St. Louis Federal Reserve (FRED), inflation rates have increased far beyond typical levels since 2020, up to 8% annually. (St. Louis Fed, 2022). With the current economic trends, sellers must now factor inflationary trends into the pricing homes, as home value can increase over time, but real value (inflation adjusted) could actually decrease over time.

Home prices are not only driven by major macro-economic factors like inflation and housing inventory and seasonality. Elena Cox from Realtor.com analyzed large sets of closed housing data to determine which endogenous factors played an important factor in determining the home prices. Among the most sought-out features listed by sellers are swimming pools, views, proximity to outdoor activities, square footage, and turnkey readiness (Cox, 2022). Analyzing these endogenous factors is nothing new, however, real estate platform giants like Zillow and Trulia offer [analysis] products directly to consumers, dubbed Zestimate (Zillow, 2022) and Estimates (Trulia, 2022), respectively. These appraisal tools offer the average home seller a forecast to help inform selling decisions and have become increasingly utilized over time.

By incorporating modern exogenous and endogenous trends, this research seeks to begin establishing a framework to alleviate the cumbersome and time-consuming process of selling a home. For this research, an ensemble model is trained on prices from the Northern Virginia real estate market and tested on the Idaho real estate market to introduce some variance. A combination of case study learnings and state-of-the-art machine learning methodologies are utilized. With the help of data from the Federal Reserve and an adjustment for future economic trends, long-short-term memory models are trained to forecast future inflationary trends.

2 Literature Review

A review of publications related to factors such as home features, inflation, and housing inventory is conducted to determine what drives increases and decreases in home prices. Prior to developing a new model, current models developed in the home pricing field are reviewed, such as neural networks and time series. This will assist in identifying what has already been built and what can be enhanced.

2.1 Factors in Pricing

2.1.1 Home Features

The price of a home is not solely determined by exogenous factors. Often a seller can estimate the value of a house based on the features of the home. According to Jafri, the following factors can affect a home's price: square footage, location, number of bathrooms, number of bedrooms and neighborhood characteristics (Jafri et al., 2019).

Yuen evaluates the effect of home features on home prices using modern machine learning techniques based on a dataset of housing in Ames, Iowa. The model leverages XGBoost for prediction and the research reports a measure of fit, or R-squared, of 0.95. According to Yuen, total square feet and overall quality (which rates the overall material and finish of the home) are the key factors for buyers (Yuen, 2020). This research implies that sellers should optimize for space and a home being turnkey ready when going to market. A limitation of this study is that many variables are dropped because they are missing values. While it is true that real world data often has missing values, dropping values could lead to a loss of information and/or permit bias to creep in (Saha, 2020). Due to the omission of potentially valuable explanatory variables, the study's results should not be taken as ground truth, but rather directional when combined with similar studies.

2.2 Inflation

The US and many other nations' economies are fundamentally dependent on the housing market. In a world where the value of the dollar is constantly changing, companies and individuals are interested in understanding how the housing market will follow. The definition of inflation can vary depending on the year of the source, but the standard modern U.S. definition focuses on increases in the consumer price index (CPI) over time (Shaban, 2019). CPI is measured by tracking pricing for a basket of common goods over time. This basket does not explicitly track owned homes (which are deemed an investment), but rental prices are included. Housing accounts for 42% of the CPI weighted expenditures and is largest expense in the US consumer economy, (Weinstock, 2022). Additionally, reported in the BLS housing and CPI report of 2022 home prices are an important indicator of the health of housing market and home affordability, so the extent that housing inflation is used as an indicator of the housing market, accurate measurement is also important (Weinstock, 2022).

Overall housing inflation can be measured in several ways. For example, a methodology can focus solely on list prices or close prices, both of which offer different information about the market. Some alternative indicators were explored in this research as possible methods to measure inflationary impact on housing prices. These include:

(1) Acquisition approach, where the purchase price of homes including additional costs such as repairs and taxes are included in calculating home cost and inflation impacts.

(2) A user-cost approach, that accounts for costs associated with the home without incorporating the potential capital gains from the sale. This approach extends the theory of an equal expected opportunity cost on the return of investment of the home.

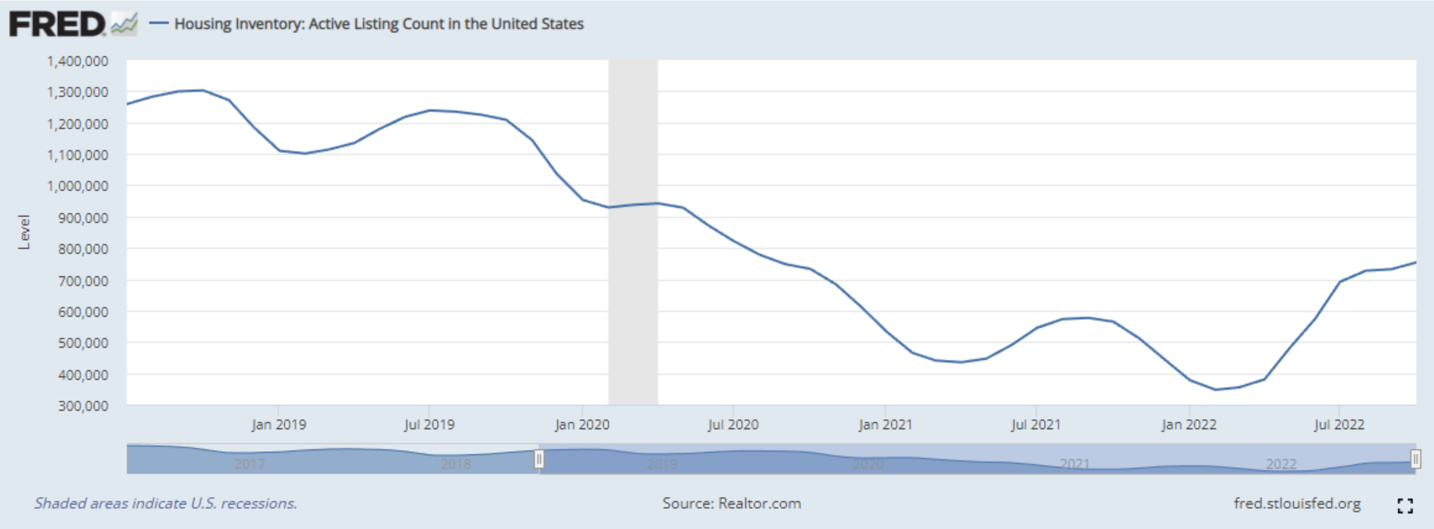
(3) A payment approach, that accounts for additional expenditures such as the mortgage payment and taxes that are required expenses for the operations of the home and living cost.

Research has shown that inflation has more impact on housing prices than housing prices have on inflation (Yu et al., 2016). With the financial crisis of 2008, the U.S. government initiated quantitative easing (QE), a strategy to inject more money into the economy. As more money begins to circulate, quantitative easing can result in increased inflationary trajectories. In periods after QE, it is important to consider how inflation will affect the sale of a home and weighing the costs and benefits of making a big sale. If, for example, inflation is rising faster than home prices, the home seller is better off selling immediately rather than waiting until summer when inflation rises and the dollar's buying power decreases.

In another study, Banking Strategist researchers look to compare housing prices and CPI. They find that from 2000 to 2020, the price of houses across the US' largest cities grew or declined at different rates (Banking Strategist, 2022). More specifically, the study finds that 95% of the US housing markets studied indicated a higher growth rate in home prices when compared to the CPI prices of homes in the same market. Therefore, the study concludes the price of homes increased at a higher rate than inflation in 2022. Furthermore, from the study when investigating the Cleveland market found that home prices did not keep pace with inflation and was a trend seen for similar major cities found in the mid-west area of the United States.

2.3 Home Inventory

Home inventory refers to the number of houses currently listed on the market as 'Active Listings'. Home inventory includes new construction and existing homes up for sale. During the COVID-19 pandemic, housing inventory plummeted due in part to government mandated mortgage forbearance. Even if homeowners could not make mortgage payments, the government allowed the deferment of the payment. This type of relief is not granted to homeowners under normal economic circumstances. Anberg et al. (2021) explains that under normal circumstances, mortgage borrowers with equity are not easily able to access forbearance or other mortgage relief that would allow them to avoid a home sale if they are unable to make their mortgage payment However, during the COVID-19 crisis, mortgage payment relief for borrowers with government-backed mortgages has been widely available and readily accessible. Under the provisions of the CARES Act, borrowers with a pandemic-related financial hardship and a mortgage from Fannie Mae, Freddie Mac, or government agencies such as FHA are entitled to up to 12 months of forbearance, which means that borrowers can defer their mortgage payments for up to a year (Anenberg et al., 2021). This constraint on supply coupled with demand for housing caused housing prices to boom. An overview of the housing inventory in the United



States from July, 2018 to October, 2022 is shown in Figure 1 (FRED, 2022).

Figure 1: Visualization of Records by Year Shows US housing inventory between July 2018 and October 2022.

2.4 House Pricing Models

2.4.1 Neural Networks

The housing market and social-economic metrics have been a prominent research area supported by many economic research literatures, especially during times of financial hardships. Tools to measure and predict volatility for home prices, observing seasonal trends for housing demands and supply as well as optimal housing consumption are key areas of interest that support the basis of the research presented in this paper. Machine learning models are tools that can be used to identify patterns in large datasets and output prediction results in new cases. Modern researchers have developed state-of-the-art algorithms that can be utilized and applied to new prediction problems.

Algorithms and methods to predict home prices are areas that contain many existing techniques that have been researched in the past. Deep Neural Networks (DNN) are known to be complex but effective for learning and identifying patterns and relationships given enough training data. There exist many types of DNNs that can be utilized output a prediction however each have different strengths and weaknesses. Convolutional Neural Networks (CNN) are great for learning feature representation and image processing but require large amounts of training data. Recurrent Neural Networks (RNN) introduce the ability to process, store and learn information that follows a sequential pattern. However, the internal memory is limited to storing the information in a single layer in the network. Originally developed by Hochreiter, Long Short-Term Memory techniques (LSTM) have gained popularity recently and have been shown to be resourceful for time series forecasting (Hochreiter et al., 1997). LSTM models advance where RNNs fall short by storing information in multiple layers enabling a longer time interval to capture information to improve accuracy in retrieving distant related memory cells.

2.4.2 Time Series

Additional methods have been proposed and implemented focused on home price predictions on a well-known dataset referred to Ames Iowa Housing Dataset (De Cock, 2011). This is a Kaggle competition dataset that is utilized to provide an open challenge to any participant in predicting the price of homes in Ames Iowa. The benefit of utilizing datasets such as this Ames Iowa Kaggle Dataset is the ability to compare methods in a structure and comparable format while enabling researching a baseline level for initial data preprocessing and model selection and model deployment. Viktorovich et al. (2018) implemented an advanced regression algorithm assembling almost 50 different machine learning models such as residual regressor, logit transform and other DNNs taking 18th place (top 1%) in the Kaggle house prediction competition (Viktorovich et al., 2018). A notable disadvantage to this approach was the increase in computation resources needed to perform the predictions. Similarly, in another solution on this dataset, C. Xiangqin (2017) assigned specific algorithms to each feature whereas the number of models utilized increased there was a similar increase demonstrated in the prediction accuracy indicating an advantage in utilizing multilevel algorithms (Chen et al., 2017). However, an abundance of time and resources are limited in selecting and experimenting with the necessary models by features to further optimize home price predictions.

3 Methods

3.1 Data Collection

The key datasets studied to evaluate the proposed approach include two existing datasets from Kaggle for home price prediction, and economic factors, pulled via direct API connection from the St. Louis Federal Reserve website (FRED).

3.1.1 NOVA Dataset

The NOVA Home Price dataset is the primary dataset utilized to train housing price prediction models in this research (Financial, 2020). It is a Kaggle challenge dataset originally web-scraped in 2020 for over 33,000 properties in Northern Virginia, from 2014 – 2018. The challenge participants were tasked with the goal of using the prepared training dataset to predict the home prices across Northern Virginia. The winner was evaluated using the root mean squared error (RMSE) accuracy metric on a held-out test dataset. While the dataset includes some records on or before 2014 and some records in 2019, records between 2014 and 2018 make up 99.3% of the dataset. In this research, the outer years are removed from the dataset before the analysis.

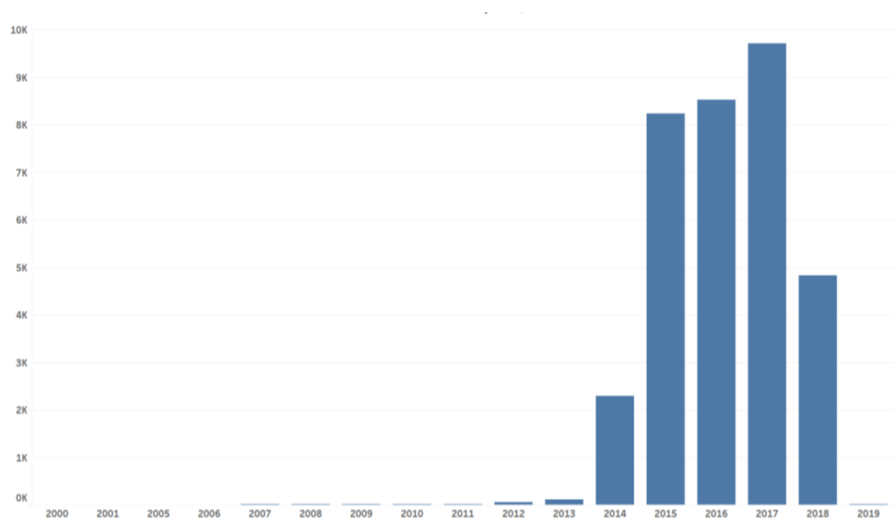


Figure 2: Number of Records by Year, NOVA Dataset

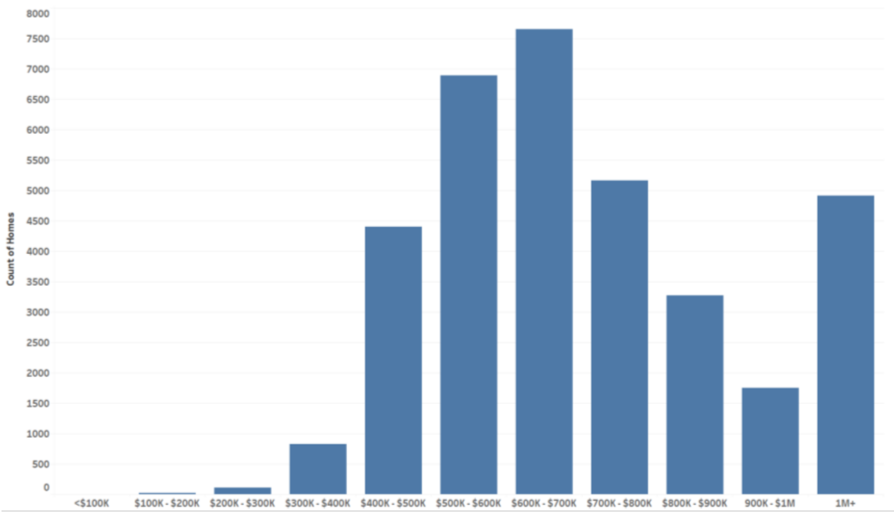


Figure 3: Distribution of Records by Home Closing Price

Figure 4 below shows the number of records by zip code in the dataset. There are 79 zip codes in total. Northern Virginia's central zip codes have fewer records, but the number of records is relatively evenly distributed after controlling for relative population. Table 1 shows the zip codes in Northern Virginia and count of the full addresses.

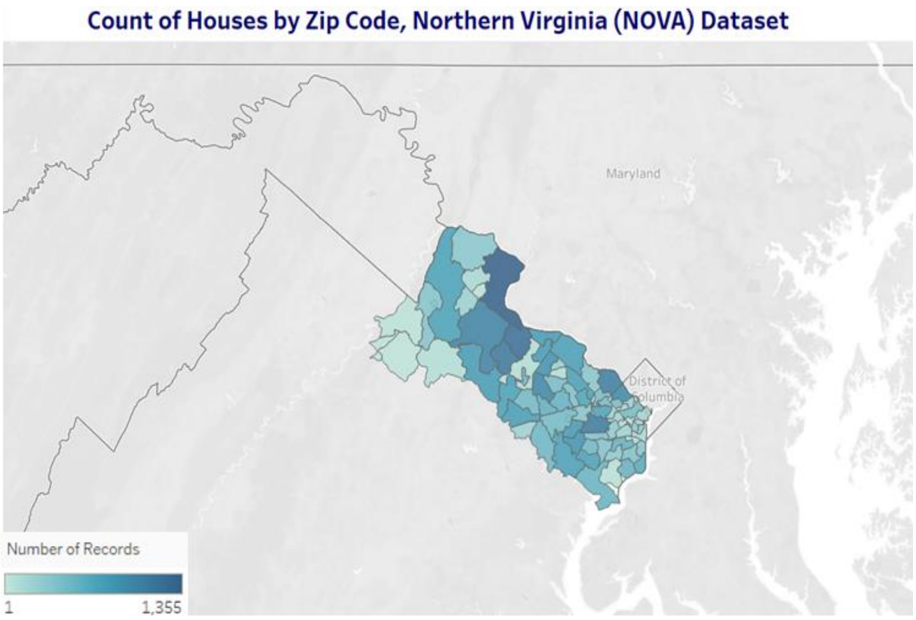


Figure 4: Record Count by Virginia Zip Code

Zip Code	Count of Full Address
20176	1355
20147	1224
20148	1130
22003	1095
22101	1072
20175	1029
20171	981
22207	955
20170	850

Table 1: Zip Codes with Highest Number of Full Addresses

3.1.2 Idaho Scraped Housing Dataset

To validate the models, data is scraped from Zillow in addition to the above datasets. Since Virginia is a relatively wealthy state, with a 2020 median income of $81,947, it is important to test model performance on a state with lower median income. Idaho's 2020 median income was $66,499, which is roughly 23% lower than Virginia's (Fred, 2022). A sample of 16 houses are scraped at random from Zillow.com and their features are passed into the pre-trained model.

Interestingly, some of Zillow's Zestimate predictions have high variances from list prices in Idaho. The random sample shows that the average deviation of list price from Zestimate is roughly $28,000, even with some houses being priced exactly as Zestimate. This may imply that Idaho may be a more volatile market, but more research would need to be done to make this claim.

3.2 Forecasting Approach

For predicting the optimal time to sell a home, a net present value (NPV) calculation is needed to adjust value over time. A long-short term memory (LSTM) framework is utilized to predict several time periods forward for inflation and housing index growth. This allows for proper adjustment of home value based on inflationary trends. Given the high variance of inflationary trends over the past decade, traditional time series models like ARIMA perform inadequately for predicting future trends. By feeding previous predictions back into the model for future predictions (called recurrence), LSTMs can offer powerful forecasts.

Before LSTM, recurrent neural networks (RNN) were one of the deep learning frameworks employed for time series modeling. LSTM helps mitigate one of RNN's biggest problems: RNNs tend to suffer from exploding and vanishing gradients. As epochs are repeated, gradient error compound gradients that quickly get very large or very small. The LSTM architecture greatly mitigates the problem of exploding and vanishing gradients associated with traditional RNNs. In order to predict the next value in a series, both LSTM and RNN use inputs of past predictions. The big distinction between the two is that LSTM offers a much more sophisticated approach, with a series of gates that "squash" the data, "forget" the point or move it on to the next gate.

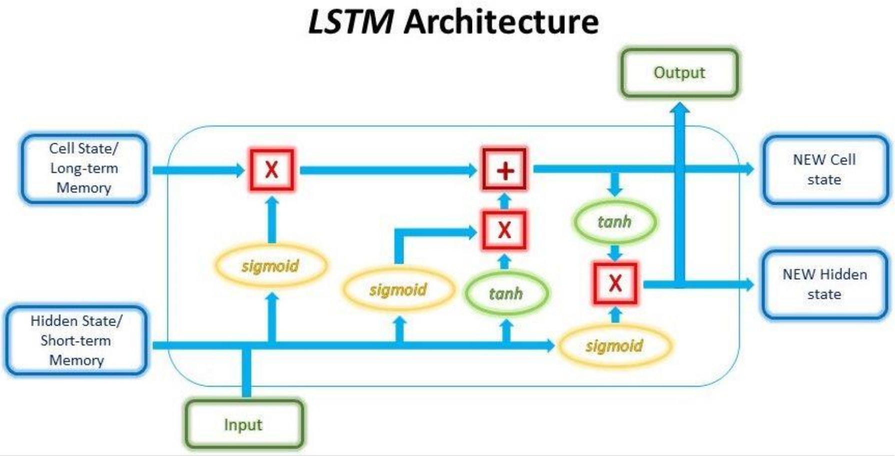


Figure 5: LSTM Architecture with Three Main Elements: An Input Gate, a Forget Gate, and an Output Gate (Loye, 2019)

The LSTM architecture used in this paper includes a 100-node initial layer, a 100-node hidden layer, both using a rectified linear activation function (ReLU) and an output layer using stochastic gradient descent as the optimized (Adam). The loss that is being optimized is mean squared error.

Two series are predicted using the architecture above: inflationary trends via consumer price index (CPI) (Fred, 2022) and localized Virginia housing trends using the Zillow Home Value Index (ZHVI) (Fred, 2022). Both series are automatically pulled from FRED utilizing a custom-built API connection. Below are the temporal characteristics of the series, along with LSTM predictions.

Zillow Home Value Index (ZHVI) is one of Zillow's data products. The ZHVI is a smoothed, seasonally adjusted measure of the typical home value and market changes across a given region and housing type. It reflects the typical value for homes in the 35th to 65th percentile range (Zillow Housing Data, 2022). Since Zillow has a wealth of data on home sales across the country, the ZHVI is used as an indicator of house price trends over time.

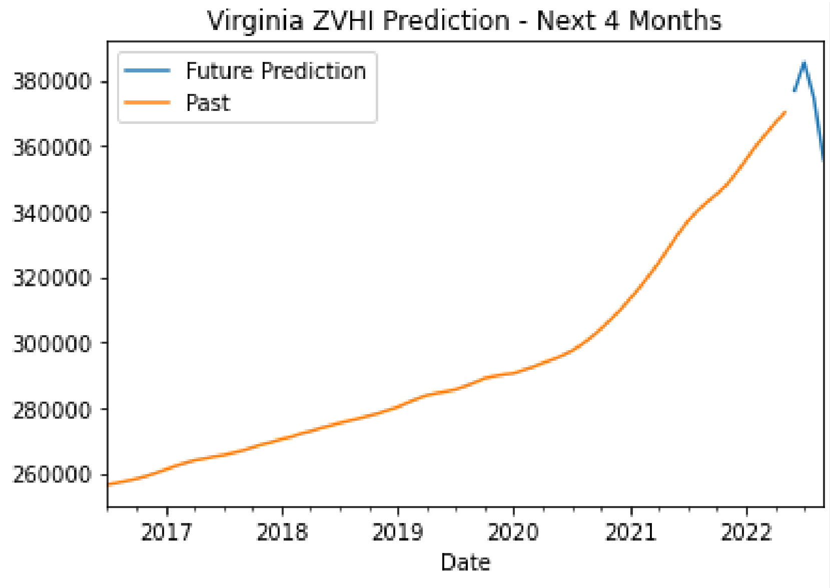


Figure 6: LSTM prediction of Virginia's Zillow Home Price Index (Fred, 2022) shows an immediate upward trend, followed by a slight correction in the near term. As of June, 2022

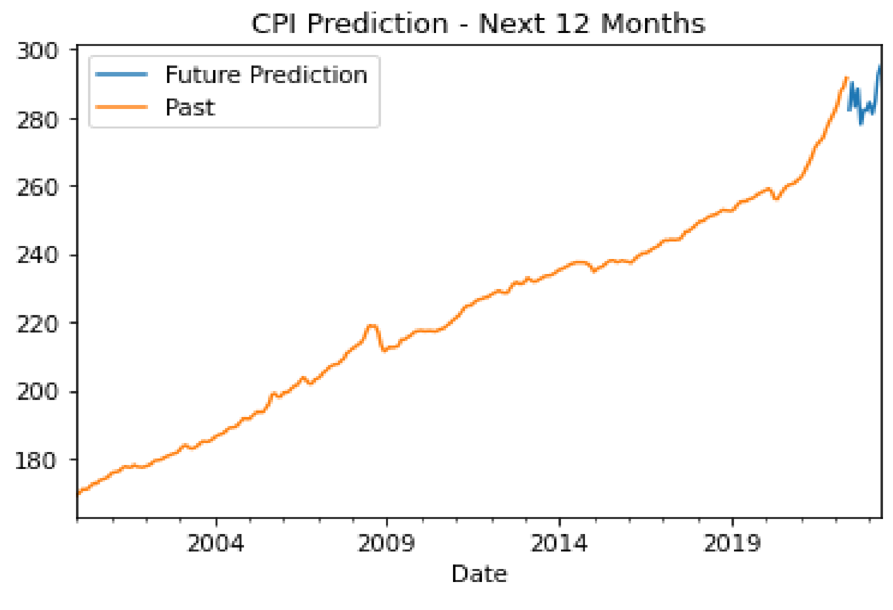


Figure 7: LSTM prediction of US CPI (Fred, 2022) shows a minor, short-term correction, followed by a steady increase in overall inflationary trends. As of June, 2022

3.3 Regression Problem – Understanding the Data

To understand the optimal time to sell, an indication of future economic trends is needed. This is answered above. These economic trends offer little value in this context without home prices to adjust. First, principal components analysis (PCA) is employed to understand which features play the largest role in explaining variance among home prices. Common intuition implies features like total square footage, number of bedrooms and bathrooms, and number of garage spaces all play exceptionally large roles in close price variance among houses. Neighborhoods also play a significant role in determining house prices. For example – Vienna, one of the richest cities in Virginia, with a median household income of $161K as of 2019 (DATAUSA, 2022), has a 0.19 correlation with home close price. Below is a correlation plot of the top 13 variables correlating with home close price, and each other. Some variables have high correlations with each other, but there are few correlations above 50%, implying a low degree of multicollinearity in the data.

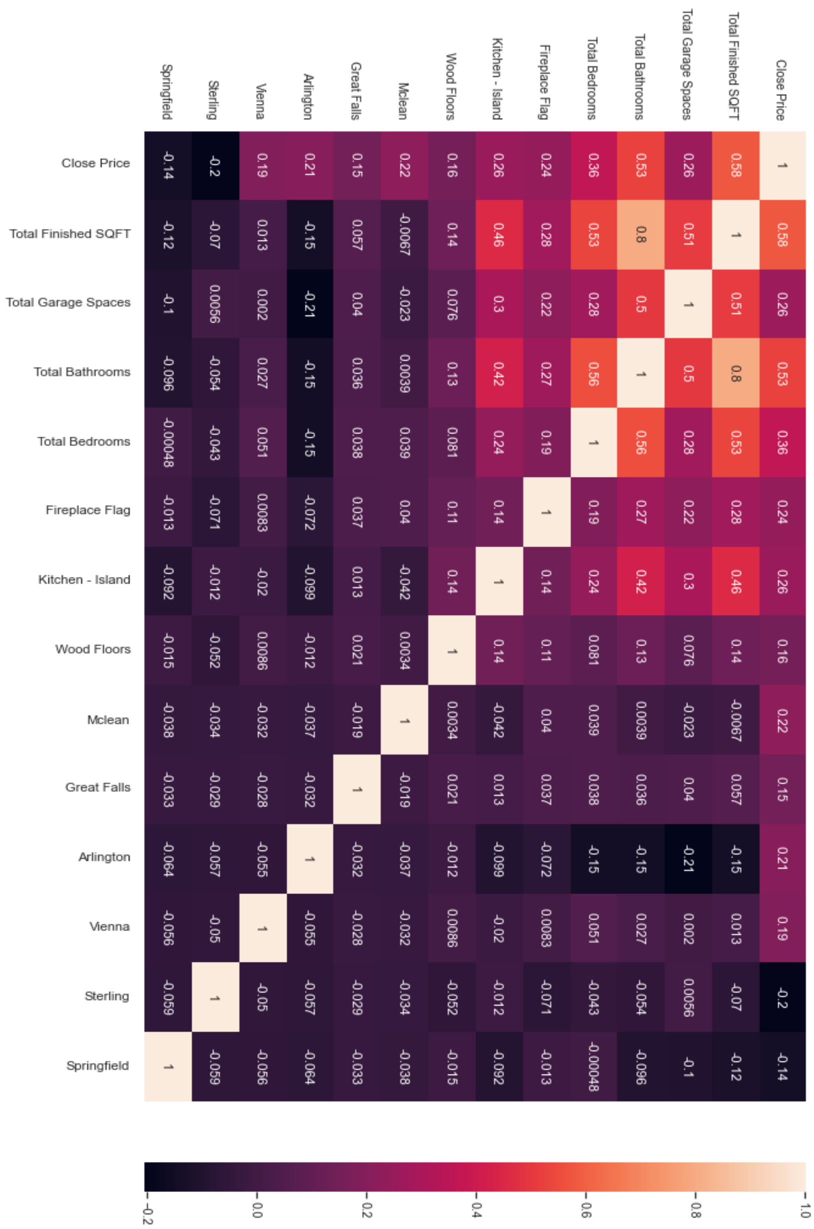


Figure 8: Correlation plot of top features (as determined by PCA) and their correlation with each other

Features	Description
FinishedSqft	Finished square footage is the completed square footage for the house which includes areas of first floor and second floor. This feature does not cover the basement square footage.
GarageSpaces	The number of cars can be parked. 3 represents 3 cars can be parked in the garage.
Bathroom	Total number of bathrooms available in the house.
Bedrooms	Total number of bedrooms available in the house.
Fireplace	Total number of fireplaces available in the house.
KitchenIsland	Categorical variable whether the house has a kitchen island or not.
Woodfloors	Categorical variable whether the house has wooden floors.
TaxAssessed	The value of the property per tax assessment.
HOA	If the house is part of home owners association (HOA).
Neighborhood	Controls for the following neighborhoods:
	Mclean', 'Great Falls', 'Arlington', 'Vienna',
	'Sterling', 'Springfield' and 'Purcellville'
FinishedBasement	If the basement is finished.
PhotoCount	Total number of photos available for the house.
ListPrice	What the house price has been listed for by the sellers.
RatioAbovetoTotal	The ratio is calculated as total basement square footage divided by the total square footage above basement.
Zestimate	The estimate Zillow machine learning model for the current value of the property.

Table 2: Feature Descriptions

Multicollinearity is further tested using variance inflation factor analysis (VIF). VIF is the ratio of overall model variance compared to the variance of a model including only a single predictor variable. Stated differently, VIF is a way to test which predictor variables account for the highest model variance. Predictor variables with high VIF indicate high collinearity with other variables in the model. The below table calculates VIF for the key predictor variables outlined in the PCA and heatmap above, as well as some other potentially important ones. There is no mathematical rule for what constitutes a high or low VIF – rather, there are sets of directional heuristics that can vary by dataset and model. Common heuristics are also noted in the table below. While the VIF of some key variables is above the 10+ heuristic, these features were tested across train, test and validation data.

VIF Range	Heuristic
1.0 – 2.5	Low Multicollinearity
2.5 – 5.0	Moderate Multicollinearity
5.0 – 10.0	High Multicollinearity
10.0+	Very High Multicollinearity

Table 3: Common (but not mathematical rules) Heuristics for VIF Ranges

Predictor Variable	VIF
Distinct VA Towns	1.1 – 1.41
Structural Features (Floors, Island, etc.)	2.0 – 3.9
Fireplace	10.7
Photo Count	13.5
All Square Footage Features	17.8 – 24.5
Total Bedrooms & Bathrooms	33.8 – 40.2
Tax Assessed Value	49.9

Table 4: VIF by Features or Sets of Features

3.4 Regression Problem – Machine Learning

The approach chosen to better understand and predict home close prices is an ensemble of common regression algorithms. Model assembling, or combining multiple models to triangulate a prediction, has been shown to deliver superior results to single model predictions (Mendes-Moreira et al., 2012). Before the final prediction is made, a pipeline of multiple models is built, and each model is evaluated on two key metrics: mean absolute error (MAE) and root mean squared error (RMSE). Two "No Skill" predictions are created to use as a baseline for evaluating advanced statistical techniques: a simple prediction of median and mean housing price across the dataset. Finally, an ensemble of XGBoost and Random Forest are equally weighted and used to predict home close prices. Since the dataset used was from 2018, the close prices were adjusted for 2022 using a scalar calculated from comparing Q12022 Virginia median home sale prices to Q12018 prices.

	RMSE	**MAE**
Simple Median	$185,236	$142,389
Simple Mean	$183,257	$143,957
Linear Regression	$84,650	$56,099
XGBoost	$67,929	$47,419
Random Forest	$67,571	$47,443

Table 5: Performance by Model in Predicting Virginia Home Close Prices

4 Results

The ensemble approach to predicting home close prices performs relatively well, but there is room for improvement. Despite the recent boom in housing prices and inflation, the model, on average is roughly $47,000 off from actuals. While more work is needed to improve performance of underlying temporal forecasts and home price predictions, the overall process performs as expected and compares nominal (non-inflation adjusted) and real gains to selling a home and can recommend optimal selling time to maximize gains. In the below example on a $500,000 home, the algorithm predicts December and January are the optimal times to sell due to rising inflation and likely falling home indices – these months are predicted to be the highest value months before inflation and falling home prices lower the real gains on sale.

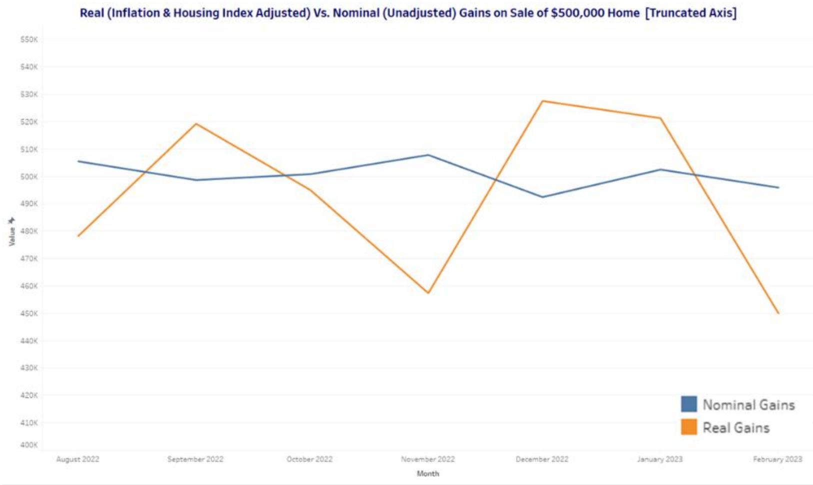


Figure 9: Real and Nominal Gains on a $500,000 Home as Predicted by Forecast

To further evaluate model performance, the close price prediction algorithm is tested on a random subset of scraped Idaho housing data. The error metrics are relatively in-line with Virginia's error metrics, implying that while more work is needed to improve model performance, the model currently has relatively low variance, even when introduced to a radically different state. Idaho results are before adjusting for neighborhood effects, which seemed to play an important role when predicting Virginia close prices. Additionally, it is important to note

	RMSE	MAE
Idaho Predictions	$86,638	$77,020

Table 6: RMSE and MAE of Predictions on Random Idaho Homes

Additionally, the algorithm can offer a series of "next best actions" based on location and housing type to increase the value of a home. Below are the top 3 in, for example, the city of Vienna, Virginia. Based on the value of the home over time above, a home seller can choose to engage in these next best actions or go to market immediately.

Next Best Action	Estimated Incremental
Additional Bathroom	$38K
Change to Wooden Floors	$36K
Addition of Kitchen Island	$3.5K

Table 7: Estimated Incremental Contribution to House Price by House Feature

5 Discussion

The framework outlined in this study uses state-of-the-art machine learning methods to first predict the value of a home's closing price in the current time period, then applies long-term economic forecasting to determine that sale's value will change over time in real, inflation-adjusted dollars. The results of this analysis show that knowing the value of a home at a particular time, while very important, is not sufficient, especially in times of rapid inflation. By using modern time series forecasting techniques to predict future inflationary trends and housing price growth, it is possible to show the real value of a home sale over time, and thus predict the optimal time to sell. It is important to note that this paper lays the foundation and framework for this type of analysis, but more work is needed to build upon these results.

The implications for this type of analysis are vast, primarily aiding independent home sellers. In the future, a framework like this may help independent home sellers forgo the need for a listing agent or initial appraisal, potentially saving thousands of dollars on home sale. The framework could also be used in conjunction with major real estate companies' platforms like Zillow, RedFin, and Trulia to drive additional efficiencies.

5.1 Ethics

From a data privacy standpoint, as this data is publicly available, this research does not use personally identifiable information (PII). There are no data privacy concerns, but generalizability is a concern. Since the training data for this research uses specifically home close prices from Northern Virginia, a relatively affluent geographical area, it is not representative of the entire US. The model performs well on disparate data scraped from Idaho, but not as well as on a holdout set from Virginia. This implies that there may be model performance disparity when applied to different regions of the US. More work must be done around data collection and model construction to ensure that this framework is scalable and applicable across the US. Additionally, macroeconomic trends like interest rates are not in the scope of this research and will be included in future iterations.

5.2 Future Research

Arguably the most important would be to incorporate feedback from subject matter experts, both in the fields of time series forecasting, and home close price prediction. More research A next step in terms of platform would be to analyze successful home sales and their photo content, offering sellers more advanced tools for selling a home as well as real-time photo feedback optimize close price. Further, a web application could be developed that would offer sellers a single platform to see all the discussed insights and future enhancements.

As this is preliminary work, these results should be viewed as the groundwork for similar analysis going forward. Home valuation and trend prediction are two very disparate fields, each with their sets of complexities and nuances. For example, with home selling, there are immeasurable factors like home seller / home buyer bias, "love letters" exchanged prior to home purchase, and corporate cash-only offers winning over higher, but leveraged offers. Controlling for these factors may be important but would require extensive further research and consultation.

6 Conclusion

Currently, home sellers can get an idea of the housing market by using existing platforms for home sales and appraisals, but these platforms cannot forecast inflationary trends nor suggest the best time to sell a home. As a result, home sellers must conduct extensive research themselves or hire a third party to do it for them at a cost. The objective of this research was to build the foundation for a way to reduce friction during the home selling process by providing homeowners with a simple recommendation on price and time to sell. The study aimed to predict the sale price of a home, determine the optimal selling period, and recommend ways to increase its value. This is achieved using a dataset of features from 34,973 Northern Virginia home listings as well as a validation dataset scraped directly from Zillow. Using machine learning techniques, an ensemble of models is trained and predicts the sale price of a house within $47,000 of the actual closing price. The model is then tested on disparate Idaho data scraped from Zillow to ensure low model variance. Despite Northern Virginia's vast differences in income, demographics, and unique home features, the model provides relatively accurate results when applied to Idaho home listings. Based on the findings of this study, it is likely that a home's features and inflationary trends play an important role in predicting home prices. Due to rising inflation and falling home indices, every seller has an ideal time to sell a home before inflation reduces the real gain on sale. There is potential for further exploration in this area such as the inclusion of interest rates, more granular controlling of neighborhood effects and expansion to broader geographic areas of the US. Despite its limitations, this study offers real estate valuation professionals an initial template for utilizing new sources of data to improve existing models of house price prediction.

References

1. Acciani, C., Fucilli, V., & Sardaro, R. (2011). Data mining in real estate appraisal: A model tree and multivariate adaptive regression spline approach1. Aestimum, (58), 27-45. doi:10.13128/Aestimum-9560
2. Carlucci, M., Grigoriadis, E., Venanzoni, G., & Salvati, L. (2018). Crisis-driven changes in construction patterns: Evidence from building permits in a mediterranean city. Housing Studies, 33(8), 1151-1174. doi:10.1080/02673037.2017.1421910
3. Flynn, T. F. (2010). Development of a forecasting model to predict the downturn and upturn of a real estate market in the inland empire. Dissertation.com.
4. Mayer, Y. G., & Nothaft, F. E. (2021). Appraisal overvaluation: Evidence of price adjustment bias in sales Comparisons. Real Estate Economics. https://doi.org/10.1111/1540-6229.12351
5. Das, P., Smith, P., & Gallimore, P. (2017). Pricing Extreme Attributes in Commercial Real Estate: the Case of Hotel Transactions. The Journal of Real Estate Finance and Economics, 57(2), 264–296. https://doi.org/10.1007/s11146-017-9621-4
6. Hupalo, M. (2021). The politics of Zestimate: Merging technology and real estate industries. The Politics of Zestimate: Merging Technology and Real Estate Industries, 3,215–219.https://radicalhousingjournal.org/wp-content/uploads/2021/12/RHJ_Issue-3.2_14_Update_Hupalo_215-219.pdf
7. The Fed and the Dual Mandate: In Plain English. (2022, June 28). Federal Reserve Bank of St. Louis. Retrieved June 20, 2022, from https://www.stlouisfed.org/in-plain-english/the-fed-and-the-dual-mandate#:%7E:text=The%20Federal%20Reserve%20System%20has,other%20words%2C%20conducting%20monetary%20policy
8. Ballentine, C. B., & Wells, C. W. (2021, November 3). *'Don't Buy Zillow Homes': A Tale of Failure, Mistrust and Hot Housing Markets*. Bloomberg. https://www.bloomberg.com/news/articles/2021-11-03/why-is-zillow-selling-7000-homes-tale-of-failure-mistrust-hot-housing-market
9. D. Schneider, "Machine learning predicts home prices," in IEEE Spectrum, vol. 56, no. 1, pp. 42-43, Jan. 2019, doi: 10.1109/MSPEC.2019.8594795.
10. The accuracy of home owners' estimates of house value, Journal of Housing Economics,Volume 2, Issue 4, 1992, Pages 339-357, ISSN 1051-1377, https://doi.org/10.1016/1051-1377(92)90008-E. https://www.sciencedirect.com/science/article/pii/105113779290008E
11. Mass appraisal: An introduction to multiple regression analysis for real estate valuation. Journal of Real Estate Practice and Education, 7(1):65–77, 2004.
12. https://www.wsj.com/podcasts/the-journal/how-zillow-failed-at-flipping-homes/c655e35f-54d6-4e35-b7a7-142bc616c4d6
13. https://en.wikipedia.org/wiki/Long_short-term_memory#/media/File:LSTM_cell.svg
14. O. M. T. &. B. S. Poursaeed, "Vision-based real estate price estimation," Machine Vision and Applications, no. 29, pp. 667-676, 2018.
15. Zheng Yuan & Yuhong Yang (2005) Combining Linear Regression Models, Journal of the American Statistical Association, 100:472, 1202-1214, DOI: 10.1198/016214505000000088
16. Chen, X., Wei, L., & Xu, J. (2017). House Price Prediction Using LSTM. https: doi.org/10.48550/arXiv.1709.08432
17. Cheri, X. (2021). Optimizations of Training Dataset in House Price Estimation. 2nd International Conference on Big Data Economy and Information Management (BDEIM), 21, doi: 10.1109/BDEIM55082.2021.00047.
18. De Cock, D. (2011). Ames, iowa: alternative to the boston housing data as an end of semester regression project. Journal of Statistics Education, 3(9). DOI:10.1080/10691898.2011.11889627
19. Financial, N. (2020). Kaggle Community Prediction Competiton. https://www.kaggle.com/competitions/novahomeprice/data
20. Hochreiter, S., & & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780. doi: https://doi.org/10.1162/neco.1997.9.8.1735
21. Sebt, M., Ghasemi, S., & Mehrkian, S. (2021). Predicting the number of customer transactions using stacked LSTM recurrent neural networks. Soc. Netw. Anal. Min, 86. https://doi.org/10.1007/s13278-021-00805-4
22. Viktorovich, P. A., Aleksandrovich, P. V., Leopoldovich, K. I., & Vasilevna, P. I. (2018). Prices of the Houses Using Regression Methods of Machine Learning. 2018 3rd Russian-Pacific Conference on Computer Technology and Applications (RPC), 1-5. doi: 10.1109/RPC.2018.8482191.

23. Yu, L., Jiao, C., Xin, H., Wang, Y., & Wang, K. (2018). Prediction on housing price based on deep learning. International Journal of Computer and Information Engineering, 12(2), 90-99. doi.org/10.5281/zenodo.1315879
24. Loye, Gabriel. (2019) Long Short-Term Memory: From Zero to Hero with PyTorch [E-Resource]. FloydHub. Retrieved Friday, July 15, 2022, from https://blog.floydhub.com/long-short-term-memory-from-zero-to-hero-with-pytorch/
25. FRED (2022), Consumer Price Index for All Urban Consumers: All Items in U.S. City Average (CPIAUCSL). St. Louis, MO: Federal Reserve Bank of St. Louis. [Software, E-Resource]. Retrieved Friday, July 15, 2022, from https://fred.stlouisfed.org/series/CPIAUCSL.
26. FRED (2022), Zillow Home Value Index (ZHVI) for All Homes Including Single-Family Residences, Condos, and CO-OPs in Virginia (VAUCSFRCONDOSMSAMID). St. Louis, MO: Federal Reserve Bank of St. Louis. [Software, E-Resource]. Retrieved Friday, July 15, 2022, from https://fred.stlouisfed.org/series/VAUCSFRCONDOSMSAMID.
27. DATAUSA (2022), Vienna, VA. Data USA [E-Resource] Retrieved Friday, July 15, 2022, from https://datausa.io/profile/geo/vienna-va.
28. Zillow (2022), What is a Zestimate? [E-Resource] Retrieved Friday, July 15, 2022, from https://www.zillow.com/z/zestimate/
29. Trulia (2022), What is a Trulia Estimate? [E-Resource] Retrieved Friday, July 15, 2022, from https://www.trulia.com/info/trulia-estimates/
30. Google Trends (2022), Interest Over Time for Search Term "Zillow". [E-Resource]. Retrieved Friday, July 15, 2022 from https://trends.google.com/trends/explore?date=all&geo=US&q=zillow
31. Cox, E. (2022, March 22). Attention, Sellers: These Are the Features Buyers Demand in Your Home—and the Ones They Can Live Without. Real Estate News & Insights | Realtor.Com®. https://www.realtor.com/news/trends/the-top-amenities-homebuyers-are-looking-for-ahead-of-spring-buying-season/
32. FRED (2022), Housing Inventory: Active Listing Count in the United States (ACTLISCOUUS). St. Louis, MO: Federal Reserve Bank of St. Louis. [Software, E-Resource]. Retrieved Friday, July 15, 2022, from https://fred.stlouisfed.org/series/ACTLISCOUUS.
33. Housing Market Indicators Monthly Update. (2022, June 30). US Department of Housing and Urban Development. Retrieved June 25, 2022, from https://www.huduser.gov/portal/ushmc/home.html.
34. Jafari, A., & Akhavian, R. (2019). Driving forces for the US residential housing price: a predictive analysis. Built Environment Project and Asset Management, 9(4), 515–529. https://doi.org/10.1108/bepam-07-2018-0100
35. Leombruno, M., Piazzesi, M., Schneider, M., & Rogers, C. (2020). Inflation and the Price of Real Assets. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3535330
36. Ersoz, F., Ersoz, T., & Soydan, M. (2018). Research on Factors Affecting Real Estate Values by Data Mining. Baltic Journal of Real Estate Economics and Construction Management, 6(1), 220–239. https://doi.org/10.2478/bjreecm-2018-0017
37. Cohen, V., & Karpavičiūtė, L. (2017). The analysis of the determinants of housing prices. Independent Journal of Management & Production, 8(1), 49–63. https://doi.org/10.14807/ijmp.v8i1.521
38. Yu, H., & Huang, Y. (2016). Regional heterogeneity and the trans-regional interaction of housing prices and inflation: Evidence from China's 35 major cities. Urban Studies, 53(16), 3472–3492. https://doi.org/10.1177/0042098015617882
39. Mendes-Moreira, J., Soares, C., Jorge, A. M., & Sousa, J. F. D. (2012). Ensemble approaches for regression. ACM Computing Surveys, 45(1), 1–40. https://doi.org/10.1145/2379776.2379786
40. Dropping Missing Values? You Probably Shouldn't. (2020, February 17). Jarmos. https://jarmos.netlify.app/posts/few-reasons-to-not-drop-missing-values/
41. Goodman, J. L., & Ittner, J. B. (1992). The accuracy of home owners' estimates of house value. Journal of Housing Economics, 2(4), 339–357. https://doi.org/10.1016/1051-1377(92)90008-e
42. Benjamin, J., Guttery, R., & Sirmans, C. (2004). Mass Appraisal: An Introduction to Multiple Regression Analysis for Real Estate Valuation. Journal of Real Estate Practice and Education, 7(1), 65–77. https://doi.org/10.1080/10835547.2004.12091602
43. Shaban, O. S., Al-Attar, M., Al-hawatmah, Z., & Ali, N. N. (2019). Consumer Price Index (CPI) as a competitiveness inflation measure: Evidence from Jordan. Journal of Governance and Regulation, 8(2), 17–22. https://doi.org/10.22495/jgr_v8_i2_p2
44. Cui, F. (2020). Quantitative study on factors affecting the price of residential real estate multiple linear regression model. Journal of Physics: Conference Series, 1629(1), 012071. https://doi.org/10.1088/1742-6596/1629/1/012071

45. Zhou, Y. (2020). HousingSalePrice Prediction Using Machine Learning Algorithms. Housing Sale Price Prediction Using Machine Learning Algorithms, 1–39. https://escholarship.org/uc/item/7rr4j05s
46. Rockett, D. (2022, April 12). *Home buyers are facing yet another challenge to the real estate market: Rising interest rates*. Chicago Tribune. Retrieved July 27, 2022, from https://www.chicagotribune.com/real-estate/ct-rising-interest-rates-and-homebuying-tt-0411-20220412-uyep6ho6onhjbmduhihbugn2aq-story.html
47. Lambert, L. (2021, July 27). *Something big is about to happen in the housing market*. Fortune. Retrieved July 30, 2022, from https://fortune.com/2021/07/26/2021-housing-market-real-estate-foreclosures-mortage-forbearance/
48. Lambert, L. (2022, June 15). *The 'Housing Correction' to intensify as mortgage rates top 6% and notch the biggest jump since 1981v*. Fortune. Retrieved July 30, 2022, from https://fortune.com/2022/06/15/mortgage-rates-spike-housing-market-real-cost-to-buy-a-home-jumps-50-percent/
49. Anenberg, E., & Scharlemann, T. (n.d.). *The effect of mortgage forbearance on house prices during COVID-19*. The Fed - The Effect of Mortgage Forbearance on House Prices During COVID-19. Retrieved July 27, 2022, from https://www.federalreserve.gov/econres/notes/feds-notes/the-effect-of-mortgage-forbearance-on-house-prices-during-covid-19-20210319.htm
50. Guler, B., & Arslan, Y. (2010). Housing prices and interest rates: A theoretical analysis. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1237722
51. Duca, J. V., & Murphy, A. (2021, December 28). *Why House prices surged as the COVID-19 pandemic took hold*. Dallasfed.org. Retrieved July 28, 2022, from https://www.dallasfed.org/research/economics/2021/1228.aspx
52. Bureau, U. S. C. (2022, April 13). *Remote working, commuting time, life events all affect home buyers' decisions*. Census.gov. Retrieved August 3, 2022, from https://www.census.gov/library/stories/2021/10/zillow-and-census-bureau-data-show-pandemics-impact-on-housing-market.html
53. FRED (2022), Median Household Income in Virginia (MEHOINUSVAA646N). St. Louis, MO: Federal Reserve Bank of St. Louis. [Software, E-Resource]. Retrieved Friday, September 11, 2022, from https://fred.stlouisfed.org/series/MEHOINUSVAA646N
54. FRED (2022), Median Household Income in Idaho (MEHOINUSIDA646N). St. Louis, MO: Federal Reserve Bank of St. Louis. [Software, E-Resource]. Retrieved Friday, September 11, 2022, https://fred.stlouisfed.org/series/MEHOINUSIDA646N
55. Banking Strategist. (2022, June). Housing Prices - HPI vs CPI. Retrieved from backing strategist: https://www.bankingstrategist.com/housing-prices-hpi-vs-cpi#
56. Weinstock, L. R. (2022). Housing and the Consumer Price Index. Congressional Research Service. Retrieved from https://sgp.fas.org/crs/misc/IF12164.pdf
57. Belz, S., & Wessel, D. (2020, January 29). Explaining the inflation puzzle. Brookings. Retrieved September 24, 2022, from https://www.brookings.edu/product/explaining-the-inflation-puzzle/
58. Zillow Housing Data (2022), Housing Data [E-Resource] Retrieved September 24, 2022, from https://www.zillow.com/research/data/
59. Merriam-Webster. (n.d.). Inflation definition & meaning. Merriam-Webster. Retrieved September 20, 2022, from https://www.merriam-webster.com/dictionary/inflation